

Arnold, Karl-Heinz

Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung

Zeitschrift für Pädagogik 48 (2002) 5, S. 741-764



Quellenangabe/ Reference:

Arnold, Karl-Heinz: Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung - In: Zeitschrift für Pädagogik 48 (2002) 5, S. 741-764 - URN: urn:nbn:de:0111-opus-38595 - DOI: 10.25656/01:3859

<https://nbn-resolving.org/urn:nbn:de:0111-opus-38595>

<https://doi.org/10.25656/01:3859>

in Kooperation mit / in cooperation with:

BELTZ

<http://www.beltz.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

Inhaltsverzeichnis

Thementeil: Evolutionäre Pädagogik

Annette Scheunpflug

Evolutionäre Pädagogik. Einführung in den Thementeil 649

Alfred K. Trembl

Evolutionäre Pädagogik – Umriss eines Paradigmenwechsels 652

Karl Ernst Nipkow

Möglichkeiten und Grenzen eines evolutionären Paradigmas
in der Erziehungswissenschaft 670

Eckart Voland/Renate Voland

Erziehung in einer biologisch determinierten Welt –
Herausforderung für die Theoriebildung einer evolutionären
Pädagogik aus biologischer Perspektive 690

Nicole Becker

Perspektiven einer Rezeption neurowissenschaftlicher
Erkenntnisse in der Erziehungswissenschaft 707

Dieter Neumann

Ein Klassiker der Pädagogik in evolutionärer Perspektive:
Eduard Sprangers „Lebensformen“ im Lichte der modernen
Biologie 720

Allgemeiner Teil

Karl-Heinz Arnold

Schulentwicklung durch Rückmeldung der Lernwirksamkeit
an die Einzelschule: Möglichkeiten und Grenzen der
Schuleffizienzforschung 741

<i>Ulrich Frick/Maria Kurz-Adam/Michael Köhler</i>	
Die Ziele der stationären Jugendhilfe – eine Typologie fachlicher Ziele und Zuweisungsmuster in der Hilfeplanung des Jugendamtes	765
<i>Deutsche Forschungsgemeinschaft (DFG)</i>	
Stellungnahme zur strukturellen Stärkung der empirischen Bildungsforschung. Ausschreibung von Forschungsgruppen in der Empirischen Bildungsforschung	786
 <i>Besprechungen</i>	
<i>Heinz-Elmar Tenorth</i>	
Erhard Wiersing (Hrsg.): Humanismus und Menschenbildung. Zu Geschichte, Gegenwart und Zukunft der bildenden Begegnung der Europäer mit der Kultur der Griechen und Römer	
Notker Hammerstein: Res publica litteraria. Ausgewählte Aufsätze zur frühneuzeitlichen Bildungs-, Wissenschafts- und Universitäts- geschichte	799
 <i>Wolfgang Keim</i>	
Barbara Feller/Wolfgang Feller: Die Adolf-Hitler-Schulen. Pädagogische Provinz versus Ideologische Zuchtanstalt	804
 <i>Peter Faulstich</i>	
Frank Achtenhagen/Wolfgang Lempert (Hrsg.): Lebenslanges Lernen im Beruf. Eine Grundlegung im Kindes- und Jugendalter. 5 Bände	808
 <i>Dokumentation</i>	
Pädagogische Neuerscheinungen	813

Content

Topic: Evolutionary Pedagogics

<i>Annette Scheunpflug</i> Evolutionary Pedagogics – An introduction	649
<i>Alfred K. Tremel</i> Evolutionary Pedagogics – Outlines of a change in paradigm	652
<i>Karl Ernst Nipkow</i> Possibilities and Limits of an Evolutionary Paradigm in Educational Science	670
<i>Eckart Voland/Renate Voland</i> Education in a Biologically Determined World – A challenge for a theory of evolutionary pedagogics from a biological point of view ...	690
<i>Nicole Becker</i> Perspectives of an Integration of Neuro-Scientific Findings into Educational Science	707
<i>Dieter Neumann</i> A Classic of Pedagogics from an Evolutionary Perspective: Eduard Spranger's "Forms of Life" in the light of modern biology	720

Articles

<i>Karl-Heinz Arnold</i> School Development through Feedback on the Effectiveness of Learning at the Individual School: Possibilities and limits of school effectiveness research	741
<i>Ulrich Frick/Maria Kurz-Adam/Michael Köhler</i> The Aims of Stationary Youth Welfare Work – A typology of subject-related objectives and allocation patterns in the planning of support by the youth welfare office	765

<i>Deutsche Forschungsgemeinschaft (DFG)</i>	
Developing Empirical Educational Research. Announcement of Research Groups in Empirical Educational Research	786
Book Reviews	799
New Books	813

Karl-Heinz Arnold

Schulentwicklung durch Rückmeldung der Lernwirksamkeit an die Einzelschule: Möglichkeiten und Grenzen der Schuleffizienzforschung

Zusammenfassung: Die Schuleffizienzforschung (SER: School Effectiveness Research) bietet im Setting breit angelegter Schulleistungstudien die Möglichkeit, in fairer Weise die Lernwirksamkeit der Einzelschule zu bestimmen. Allerdings ist aufgrund des Designs der meisten Studien die Generalisierbarkeit der Ergebnisse auf der Ebene der Einzelschule so weitgehend eingeschränkt, dass eine auf die Einzelschule als Handlungseinheit fokussierte Schulentwicklung kaum aufgrund dieser empirischen Evidenz planbar und evaluierbar erscheint. Hingegen bietet unterhalb der Schulebene die lehrer- bzw. klassenbezogene sowie die jahrgangs- und fachbezogene Rückmeldung der Lernwirksamkeit wertvolle Information für die traditionellen Formen der Beratung, Supervision und Lehrerfortbildung. Oberhalb der Schulebene, d.h. auf der Ebene des Schulsystems bieten SER-Daten höchstbedeutsame Evaluationsdaten zur Entwicklung der sozialen Gerechtigkeit. Wenig beachtet wird, dass auf der Ebene der Einzelschule subtile Prozesse des Risikomanagements ablaufen, die einerseits die Effektivität des Gesamtsystems sichern, andererseits jedoch kaum in SER-Kennwerten abbildbar sind.

1. Entwicklung der Einzelschule oder Reform des Bildungssystems?

Bildungs- bzw. Schulreform scheint bereits ein historischer Begriff geworden zu sein. In den 70er- und 80er-Jahren des letzten Jahrhunderts wurde unter diesem Stichwort die pädagogisch und politisch intendierte, systembezogene Veränderung des Lehrens und Lernens in den öffentlichen Schulen bezeichnet. Schuladministrative Maßnahmen wie Lehrplanreform, Veränderung der Schulformen (Gesamtschule, Integration von behinderten Schülern in die Allgemeine Schule) und Aufbau von Unterstützungseinrichtungen für Schüler, Eltern und Lehrer (z.B. Beratungsdienste, schulinterne Lehrerfortbildung) wurden begleitet von konzeptuellen Debatten über günstigere Unterrichts- und Erziehungsmethoden (z.B. reformpädagogische Unterrichtsmethoden wie Projektunterricht; sozial-integrativer Erziehungs- und Unterrichtsstil). Die empirische Schulforschung hielt und hält sich bereit zur vergleichenden Evaluation der Innovationen und folgt damit dem klassischen Prozess-Produkt-Paradigma. Das Schulsystem als Errungenschaft einer demokratisch-humanistischen Gesellschaft sollte an deren verfassungsmäßigen Zielen gemessen und diesen zunehmend angenähert werden. Schulisch bewirkte Bildung stand in der Pflicht, einen Beitrag zur sozialen Gerechtigkeit zu leisten,

indem erhebliche Ungleichheiten in den Lebensverhältnissen durch das Erreichen gleicher Lernstände in wichtigen Lernbereichen ausgeglichen werden (vgl. Ditton 1995; Preuss-Lausitz 1997; Arnold 2001).

In den 90er-Jahren des letzten Jahrhunderts, das in Deutschland erst mit der Reichsschulkonferenz von 1920 die (Volks-)Schule zu einer ‚Veranstaltung des Staates‘ gemacht hatte, scheint mit dem Tripel ‚Autonomie der Einzelschule–Schulentwicklung–Evaluation‘ eine gewisse Abwendung von der systembezogenen Ausrichtung pädagogischer Innovation vollzogen zu werden. Einige Ergebnisse der systemvergleichenden Gesamtschulstudien – die Zwischenschulvarianz übersteigt die Varianz zwischen den Schulformen (vgl. Fend 1982, S. 289) – wurden wieder entdeckt und im Kontext einer vielleicht überschätzten administrativen Steuerungswirkung der Bildungsverwaltung als Beleg für die sicherlich nie bestrittene These aufgefasst, dass die Einzelschule eine bedeutsame Handlungseinheit darstellt. Unter Hinzunahme organisationssoziologischer und arbeitspsychologischer Konzepte wurde aus verordneter ‚Reform‘ nunmehr partizipative ‚Entwicklung‘. Der Titel der vielleicht zentralen deutschsprachigen Veröffentlichung spiegelt die neue Fokussierung – mit einem selten beachteten Fragezeichen versehen – wider: „Schulentwicklung als Entwicklung von Einzelschulen?“ (Rolff 1991). Die Begriffe ‚Schulsystem‘ oder ‚Schulformentwicklung‘ scheinen nicht benötigt zu werden. Dies mag auch einen theoretischen Grund haben: Organisationsentwicklung setzt auf das Engagement der Akteure am Arbeitsplatz, der für Lehrer die eigene Schule ist.

Allerdings hatte schon die Fend-Studie ein kaum beachtetes, gleichwohl äußerst kritisches Ergebnis erbracht: offensichtlich gibt es „differentielle Förderungswirkungen unterschiedlicher Schulformen“ (Fend 1982, S. 248). In den Vergleichsstudien von Ditton/Krecker (1995) zeigen sich gleichfalls sehr beträchtliche Effekte der Schulart. Die Ergebnisse der BiJu-Studie¹ (vgl. Köller/Baumert 1999) replizieren diese Einsicht ebenso wie die MARKUS-Studie² in Rheinland-Pfalz (vgl. Hosenfeld u.a. 2001). Insofern müsste ein institutionsbezogenes Schulentwicklungskonzept auch die Frage umfassen, ob und wie institutionelle Teilbereiche voneinander lernen können, was gleichbedeutend wäre mit der Frage, ob z.B. die hohe Haltekraft ein übertragbares, allerdings implizites Merkmal des ‚Schulprogramms Gymnasium‘ darstellt: für Schüler mit ungünstiger Erfolgsprognose ist das Gymnasium deutlich erfolgreicher als andere Schulformen. Die Varianzquelle Schularten bzw. -formen gilt jedoch traditionelles Handlungsgebiet von Bildungsreform.

1 BiJu: Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (vgl. Baumert 1997).

2 MARKUS: Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext (vgl. Helmke/Jäger 2001).

2. Entwicklung der Einzelschule und Evaluation der Einzelschulentwicklung

Schulprogrammarbeit kann wichtige Effekte für andere Gruppen als die Schüler erbringen. Möglicherweise wird die Arbeitszufriedenheit der Lehrer erhöht, indem sie mehr kollegiale Kooperation, Wertschätzung und Unterstützung erleben. Schulleiter finden über die pädagogisch-konzeptionelle Arbeit einen nicht-hierarchischen Zugang zu ihren Mitarbeitern und verbessern damit ihre Position im Kollegium. Eltern erleben die Schule als engagiert und können sich an der konzeptionellen Arbeit beteiligen. Die Schulaufsicht kann ihre kontrollierenden Aufgaben reduzieren und beratende Aufgaben ausweiten. All diese Prozesse können als *systembezogene* Verbesserungen ausgewiesen werden. Ob diese Veränderungen auf der Schulebene günstige, keine oder ungünstige Wirkungen auf die Lernentwicklung der Schüler haben, ist damit noch nicht entschieden. Eine möglicherweise vorübergehende, derzeit jedoch bestehende Problematik von Schulentwicklung wird hier sichtbar. Wenn dieses Reformkonzept im Kontext einer empirischen Wendung der Schulpolitik favorisiert wird, dann sollte auch empirische Evidenz für die gewünschten, vermuteten oder behaupteten Veränderungen beigebracht werden.

Die große Leistungsfähigkeit der international vergleichenden Bildungsforschung wird in den aufwendigen Studien der IEA (insbesondere TIMSS) und der OECD (insbesondere PISA) eindrucksvoll unter Beweis gestellt (vgl. Baumert 2001). Weniger klar ist die Frage zu beantworten, welche Evaluationsleistung die empirische Schulleistungsforschung für die Ebene der Einzelschule erbringen kann. Unter dem Begriff *School Effectiveness Research* (SER; vgl. Willms 1992; Fitz-Gibbon 1996; Scheerens/Bosker 1997) ist insbesondere in der vergangenen Dekade eine internationale Forschungstradition hervorgetreten, die eine entscheidende Verbindung ermöglicht. Im Setting großer Survey-Studien werden die Merkmale von Einzelschulen und der dort unterrichteten Schüler so umfassend und präzise erfasst, dass die Lernwirksamkeit der Einzelschule abschätzbar wird – und dies in fairer Weise (vgl. Arnold 1999a), d.h. unter expliziter Berücksichtigung jener lernrelevanten Merkmale, die von Schulen nicht oder kaum beeinflussbar sind (insbesondere die kognitive Grundfähigkeit, der Sozialstatus und das Vorwissen der Schüler). Grundsätzlich stehen damit die empirischen und statistischen Mittel (insbesondere die Mehrebenenanalyse, vgl. Goldstein 1995; Ditton 1998) bereit, die auf *fachliches Lernen* bezogene Wirksamkeit einer Einzelschule zu bestimmen und als Schuleffizienzkennwerte der Schule rückzumelden.

Hingegen ist die Erfassung und schuleffizienzbezogene Verrechnung von Merkmalen, die die *erzieherische Wirkung* von Schule abbilden, aus unterschiedlichen Gründen weitaus schwieriger und zudem grundlegend proble-

matisch. Zum einen können Konstrukte wie Lernmotivation, soziale Kompetenz oder moralisches Urteilsvermögen deutlich weniger genau als fachliche Leistung gemessen werden. Die auf relativ kleinen Gruppengrößen basierenden Kennwerte (Mittelwerte für die Schüler einer Klasse, Mittelwerte für die Parallelklassen einer Schule) werden doppelt belastet: die beträchtliche Messungenauigkeit der Erfassungsinstrumente erhöht die beträchtliche Schätzungsgenauigkeit kleiner Schulstichproben.

Das zweite Problem ist theoretischer Natur: weitaus stärker als beim schulfachbezogenen Lernen und vor allem in absichtsvoller, wenngleich selten koordinierter Weise erziehen Elternhaus und Schule zeitgleich und wirken auf die Einstellungsbildung und die sozialen Fähigkeiten der Kinder ein. Die Trennung dieser permanenten Doppelerziehung in einen elternhaus- und einen schulspezifischen Anteil erscheint weder plausibel noch schul- bzw. familienpädagogisch modellierbar.

Schuleffizienzforschung als Bestimmung der Lernwirksamkeit der Einzelschule verwendet durchweg fachliche Lernstände und gibt deshalb ein deutlich eingeschränktes Bild der Gesamtwirksamkeit der Institution. Die Wertschätzung fachbezogener Lernstände scheint jedoch insbesondere bei den Eltern so groß zu sein, dass die Frage nach einer adäquaten Abbildung der Gesamtwirkung von Schule leichthin als sophistisch abgetan wird. Gleichwohl ist sie von größter schulpädagogischer Bedeutung. Dies zeigt eine Kontroverse, die mit einem bemerkenswerten Titel überschrieben ist: „Wozu ist die Schule da?“ (vgl. Giesecke 1995).

3. Das Interesse der Bildungsadministration an der Schuleffizienzmessung

Schulen unterscheiden sich beträchtlich in den durchschnittlich erreichten Fachleistungen. Dies zeigt sich als Resultat testbasierter Schulleistungsforschung. Zensurenmittelwerte differieren weit weniger zwischen den Schulen, was eine Nebenfolge der in erheblichem Maße sozialnormorientierten, d.h. klassenbezogenen Zensierungspraxis der Lehrer ist. Allerdings besteht eine beträchtliche Korrelation zwischen dem durchschnittlichen Sozialstatus der Schülerschaft einer Schule und den Schulleistungsmittelwerten: in Schulen mit günstigerem Einzugsgebiet werden günstigere Lernstände erreicht. Eine faire Kennwertberechnung muss diese von den Schulen nicht beeinflussbaren Bedingungen berücksichtigen. Als statistisches Modell wird dazu die Regressionsanalyse verwendet: die Vergleichbarkeit der unter ungleichen Bedingungen arbeitenden Schulen wird dadurch hergestellt, dass jene Merkmale, die Vorteils- bzw. Nachteilseffekte produzieren, als Regressorvariablen in die so-

mit bedingte Schätzung (Konditionierung) von Schulmittelwerten einbezogen werden.

Konditionierte Mehrebenenmodelle der Schuleffizienzforschung basieren auf einer spezifischen Form des sozialnormorientierten Vergleichens (vgl. Helmke 2001). Der faktisch erreichte Mittelwert einer Schule wird mit jenem fiktiven Wert verglichen, der für diese Schule – unter Berücksichtigung der die Vergleichbarkeit einschränkenden Schüler- und Schulmerkmale – im Ensemble aller an der Untersuchung teilnehmenden Schulen kalkulierbar ist. Die Abweichung des realen vom vorhergesagten Schulmittelwert kann positiv oder negativ (oder nicht vorhanden) sein.

Erhebliche Abweichungen werden im Kontext von Schulentwicklungsprojekten als ‚erwartungswidrig‘ interpretiert und als Anlass bzw. modifizierendes Faktum für Schulentwicklungsstrategien genommen. Abbildung 1 zeigt den Typus von Regressionsdiagrammen, die in Deutschland z.B. als Ergebnis der LAU-Studie³ (vgl. Maritzen 2001, S. 53) oder in England als Ergebnis des National Curriculum Assessment (vgl. Arnold 1999a, S. 81ff.) zur Illustration der Nutzbarmachung von „value added measures“ (vgl. Hopkins 1996, S. 45) Verwendung finden. Gegenüber der in England bestehenden und nachhaltig von Seiten der Wissenschaft kritisierten Praxis (vgl. Goldstein/Spiegelhalter 1996), nicht-adjustierte Schulmittelwerte in sog. school performance tables zu publizieren, wird damit ein erheblicher Fortschritt erreicht.

Obgleich für Schulen in diesem Modell faire Schuleffektivitätsindikatoren gerechnet werden, ergeben sich einige Probleme. Unklar ist, ob Bildungsadministrationen in der Gesamtperspektive auf das Schulsystem von Schulen anderes erwarten können als das Erreichen (und somit das Nicht-Unterschreiten) des Erwartungswertes. Muss nicht in einer demokratischen Gesellschaft das Ziel sein, dass die eingerichteten Schulen gleichermaßen effektiv sind? Bezogen auf die negativ vom Erwartungswert abweichende Einzelschule wird jedoch eine andere Position bezogen: diese Schule soll durch ‚improvement‘ ihre Effektivität steigern. Die Fortsetzung einer Schulentwicklungspolitik, die auf die Effizienzsteigerung aller erwartungswidrig schlechten Schulen setzt, gerät in eine unlogische Konstellation, wenn zugleich für die erwartungswidrig guten Schulen ein pragmatischer Standpunkt eingenommen wird, den Hopkins (1996, S. 46) wie folgt beschreibt: „to assist effective schools to remain so“. Das Problem liegt hier in der nicht vorhandenen Standardsetzung: Schuleffizienzkennwerte basieren auf einem ‚sozialen Vergleich‘ zwischen den vorhandenen Schulen. Wenn also ‚nur‘ die erwartungswidrig schlechten Schulen ihre Effizienz steigern, dann ändert sich auch die Position

3 LAU-Studie: Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen (vgl. Lehmann/Gänsfuß/Peek 1999).

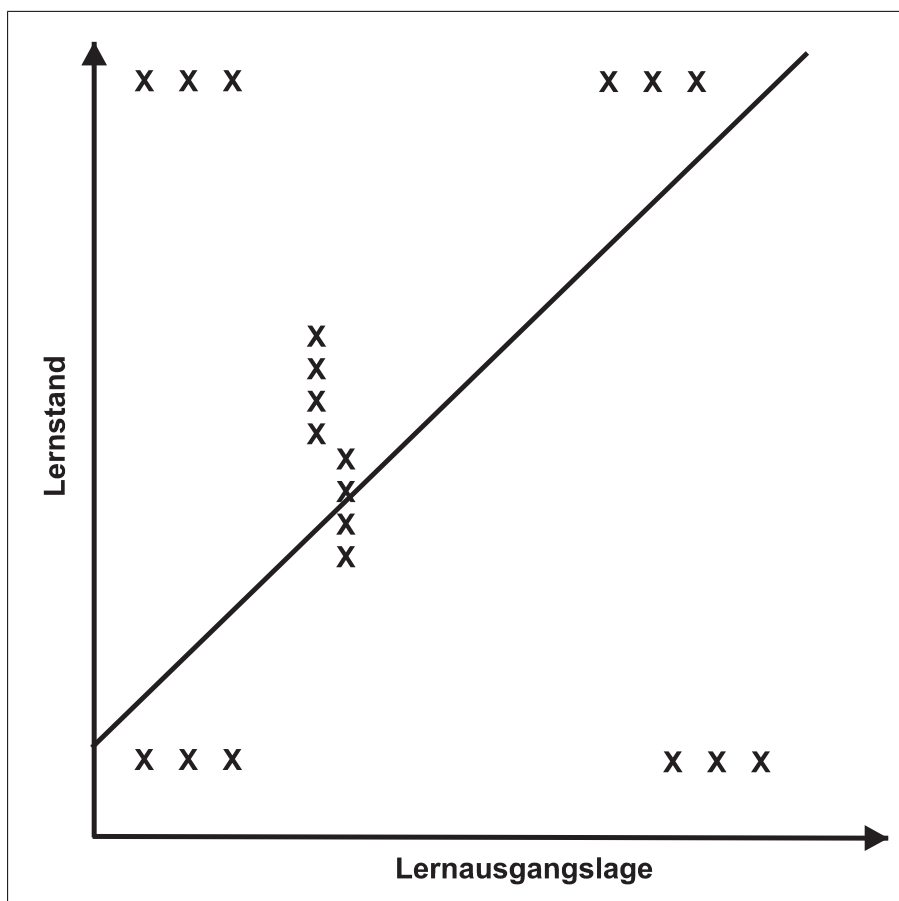


Abb. 1: Schul- bzw. Klasseneffizienz als Abweichung vom Erwartungswert (Regressionsgerade)

er erwartungswidrig guten Schulen: sie finden – ohne dass sich ihre Arbeit verändert hat – ihre Position im Diagramm herabgezogen in Richtung der Regressionsgeraden. Welche Konsequenzen Rückmeldungen dieser Art für die letztgenannte Gruppe von Schulen hat, stellt eine heikle Frage dar.

Auf der Systemebene ist diese Konstellation gänzlich anders zu bewerten. Die zentrale Botschaft von fairen Einzelschulvergleichen wird bislang offensichtlich kaum beachtet. Wird – was ein zentraler fachlicher Standard ist – das Ausmaß der Unsicherheit für die Schätzung der adjustierten Schuleffektivitätswerte mitgeteilt, so ergeben sich üblicherweise Konfidenzintervalldarstellungen wie in Abbildung 2 gezeigt.

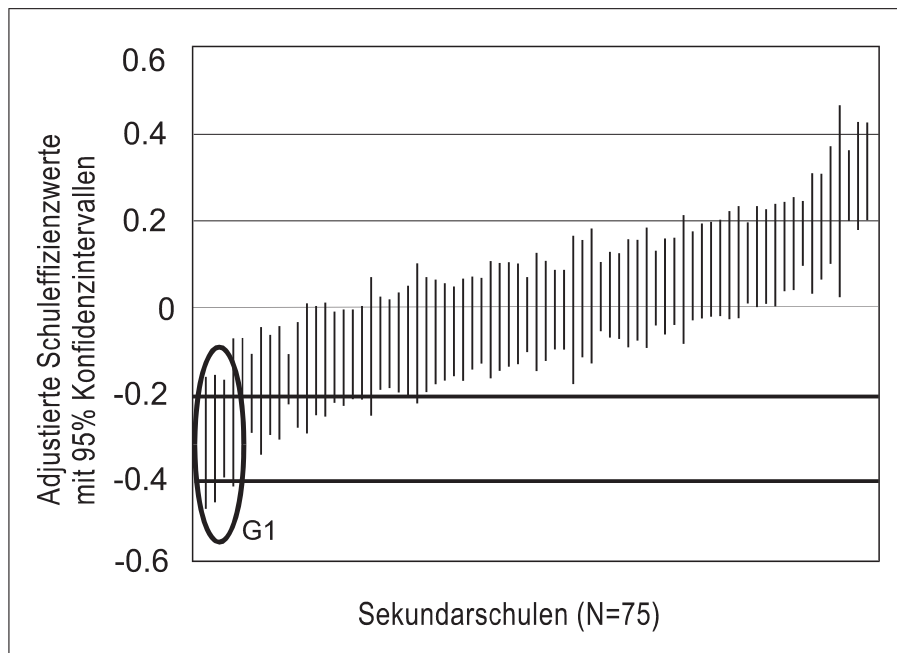


Abb. 2: Beispiel für Schuleffizienzkennwerte mit 95%-Konfidenzintervallen
(vgl. Goldstein/Spiegelhalter 1996)

Empirisches Faktum ist, dass sich in staatlich administrierten Systemen zumindest zwei Drittel der Schulen in ihrer Effektivität nicht unterscheiden, da sich die Konfidenzintervalle überlappen. Die positive Formulierung des Sachverhalts, d.h. die nicht zurückweisbare Nullhypothese lautet: Die meisten Schulen sind gleich effektiv. Ist diese Faktenlage nicht ein Zielkriterium für Bildungspolitik, die Chancengleichheit im Angebot von Bildungseinrichtungen intendiert? Auf der Ebene der Einzelschulen wird in hohem Maße Egalisierung der Lernwirksamkeit erreicht, was just jener Forderung entspricht, die Coleman u.a. (1966) wie folgt formulierten: „Thus, equality of educational opportunity implies, not merely ‘equal’ schools, but equally effective schools, whose influences will overcome the differences in starting point of children from different social groups.“

Aufgrund der ununterscheidbaren Durchschnittlichkeit der meisten Schulen lässt sich auch folgern, dass es aus der Perspektive der SER nur eine – relativ kleine – Problemgruppe von Schulen gibt. Dies sind die signifikant weniger als durchschnittlich effektiven Schulen (s. Markierung G1 in Abbildung 2). Von Maritzen und Hopkins wird implizit unterstellt, dass Begleitkonzepte bzw. Typen von Schulentwicklungsstrategien für fast alle Schulen vorgehalten

werden sollen. Dies kann nur unter der theoriefremden Voraussetzung möglich sein, dass jene Majorität der Schulen, die in dem explizit beanspruchten Modell der SER als erwartungskonform effektiv zu bezeichnen sind bzw. denen das Gegenteil nicht mit hinreichender Sicherheit nachzuweisen ist, dennoch nicht den – allerdings nicht explizierten – ‚Erwartungen‘ der Bildungsadministration entsprechen.

Diese Widersprüchlichkeit kann auch als Theorieproblem des Konzepts Schulentwicklung interpretiert werden. Die Ziele der Schulentwicklung sind letztlich bildungspolitischer Art und damit Standardsetzungen, die vorgegeben werden, auch wenn der empirische Nachweis (noch) nicht erbracht ist, dass diese von fast allen Schulen tatsächlich erreicht werden können. Das Grundkalkül der SER wird damit jedoch verlassen: statt sozialnormorientiertem Vergleichen der Einzelschulen muss eine sachnormorientierte Skala der Schuleffizienz vorgegeben werden, was erhebliche Probleme bereiten dürfte. Und ebenso kann eingewendet werden, dass dieser Vorgang der Standardsetzung typisch ist für eine von pädagogisch wohlbegründeten Idealen geleitete Bildungsreform, die Ziele setzt im Vorgriff auf für möglich Gehaltenes.

Ein weiteres Problem der schulentwicklungsbezogenen Interpretation von SER-Kennwerten betrifft die Analyseebene. Im Beitrag von Maritzen wird die Grafik zunächst korrekt als Darstellung der Effektivitätskennwerte von Schulklassen interpretiert. Eine primär auf die Schule als Handlungseinheit (und nicht die Klasse und damit den einzelnen Lehrer) ausgerichtete Schulentwicklung ist jedoch mit zwei Problemen konfrontiert.

So wird von Maritzen die Zusatzbedingung eingeführt, dass Schulentwicklungsstrategien nur für jene Schulen gelten, deren Parallelklassen gleiche SER-Kennwerte aufweisen. Aufgrund der beträchtlichen Unsicherheit in der Abschätzung von Schulklasseneffizienzwerten wird diese Konstellation häufiger vorliegen bzw. das Gegenteil trotz numerisch ungleicher Werte nicht nachweisbar sein. Die Mehrebenenzerlegung der Schuleffektivität lässt jedoch – mit der Ausnahme einer bemerkenswert hohen Parallelklassenähnlichkeit, die Luyten/de Jong (1998) in den Niederlanden fanden – zumeist folgendes Bild entstehen: die Variabilität zwischen den Klassen einer Schule übertrifft die Variabilität zwischen den Einzelschulen. Im Falle der MARKUS-Studie erklärt die Klassenzugehörigkeit ungefähr das Doppelte der Varianz, die durch Schulzugehörigkeit erklärt wird (vgl. Hosenfeld u.a. 2001). Aufgrund dieser Faktenlage ergeben sich erhebliche Zweifel, ob Schulentwicklung als Entwicklung der Einzelschule tatsächlich an der aussichtsreichsten Varianzquelle ansetzt. Begriffsneuschöpfungen wie „pädagogische Schulentwicklung“ (vgl. Bastian 1997) oder „Unterrichtsentwicklung“ (vgl. Klippert 1999) lösen das Theorieproblem durch Reintegration sehr traditioneller Konzepte der unterrichtsbezogenen Lehrerqualifizierung.

Eine zweite Folgerung betrifft die Analyseebene der Rückmeldung von SER-Kennwerten: klassenbezogene Rückmeldungen haben vermutlich einen sehr viel höheren Informationswert als Schulrückmeldungen. Evaluationsstudien wie LAU, QuaSUM⁴ und MARKUS arbeiten mit intakten Klassen als Untersuchungseinheiten und können insofern Klassenkennwerte rückmelden. PISA 2000 hingegen hat in Schulen alters- und klassenstufenbezogene Zufallsstichproben gezogen und kann deshalb ‚nur‘ Schulkennwerte rückmelden. Auch hier stellt sich nochmals die Strategiefolge der Schulentwicklung. Braucht man, um schulklassenbezogene Effizienzsteigerung zu erreichen, ein Konzept, an dem die gesamte Schule beteiligt ist?

4. Lernwirksamkeit der Einzelschule: Qualität eines sozialwissenschaftlichen Konstrukts

Die zur Schulentwicklung genutzten Rückmeldungen über SER-Kennwerte sollen nun unter diagnostischem Gesichtspunkt analysiert werden. Aus wissenschaftlicher Sicht wertvolle Information können diese den Schulen nur liefern, wenn hinreichend präzisierbar ist, was hier gemessen wird (Validität) und in welchem Maße das Gemessene Geltung besitzt für ähnliche Messgelegenheiten (Zeitstabilität), ähnliche Messinhalte (Konsistenz über Unterrichtsfächer) und Personengruppen (Parallelklassen und Jahrgänge bzw. Kohorten).

Schuleffizienz ist nicht direkt beobachtbar. Es handelt sich hier um ein Konstrukt, d.h. um eine theoriegeleitete Verknüpfung (Hierarchisch Lineare Modellierung, HLM) einzelner messbarer Merkmale (Prädiktoren auf Schüler- und Schulebene sowie Kriterien auf Schülerebene). Das SER-Modell beansprucht, die Lernwirksamkeit einer Schule unter Berücksichtigung leistungsrelevanter Merkmale höherer (z.B. Schulform), gleicher (z.B. Ausstattung der Schule) sowie niedrigerer Ebenen (z.B. sozioökonomischer und Bildungsstatus des Elternhauses) abschätzen zu können.

Die Anzahl der hierarchischen Ebenen, die als leistungsrelevante Aggregatbereiche begründet werden können, wird oberhalb der Einzelschule weitgehend durch die Verwaltungsstruktur des Schulwesens bestimmt (Schulform bzw. -art, Schulbezirk, Bundesland). Unterhalb der Ebene der Einzelschule liegen die Ebene der Schulklasse und die Ebene des einzelnen Schülers. In Abhängigkeit vom gewählten Untersuchungsdesign können jedoch weitere Ebenen unterschieden werden. Eine Querschnittuntersuchung mit Fokus auf

4 QuaSUM: Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (vgl. Lehmann u.a. 2000).

die Einzelschule unterscheidet die Ebene des Jahrgangs und darunter die Ebene der Parallelklassen im Jahrgang. Eine Längsschnittuntersuchung der Einzelschule verwendet verschiedene Messzeitpunkte auf der Ebene ‚unterhalb‘ der Ebene des Einzelschülers.

Die empirische Trennung von Schul- und Klasseneffekten ist schwierig, da Lehrereffekte über unterschiedliche Klassen und Fächer vermittelt eintreten. Für eine hinreichend präzise Separierung von Lehrereffekten müssten zumindest zwei Parallelklassen in den Schulen sowohl von der gleichen und eine weitere von einer anderen Lehrperson unterrichtet werden. Da diese Faktenlage selten gegeben ist, bleiben Schul-, Klassen- und Lehrereffekte konfundiert. Der vergleichenden Interpretation von rückgemeldeten SER-Schul- und Klassenkennwerten ist hier eine empfindliche Grenze gesetzt.

Die von Scheerens/Bosker (1997, S. 77) vorgelegte Metaanalyse der Lerneffizienz von Einzelschulen erbringt folgende Abschätzung. Wenn konditionierte Modelle gerechnet werden (net school effect), so können maximal 8 Prozent der Schulleistungsvarianz durch Merkmale von Einzelschulen erklärt werden. Werden Kontextmerkmale von Schulen mitberücksichtigt, so reduziert sich der Netto-Schuleffekt auf 4 Prozent. Die MARKUS-Studie erbringt ähnliche Befunde: der generelle Schuleffekt beträgt 3,4% der Testleistungsvarianz, bei Berücksichtigung von Kontextmerkmalen reduziert sich dieser auf 2,6% (vgl. Hosenfeld u.a. 2001). In dieser Größenordnung liegen allerdings auch die mit problematischen statistischen Modellen gerechneten Schuleffekte, die von Coleman in den 1960er-Jahren und einige Jahre danach von Jencks publiziert worden sind. Die damalige, durchaus polemisch zugespitzte Botschaft lautete: *Schools don't make a difference*.

Das methodische Gegenargument ist nochmals geprüft worden. Nach sorgfältigem Vergleich unterschiedlicher statistischer Modellierungen aktueller Daten gelangen Cheung/Keeves (1990, S. 307) zu folgender Feststellung: „[...] it has commonly been accepted that the learning conditions in schools and classrooms have little effect on student learning and that the characteristics of teachers make little contribution to the performance of the students in their classes. It is not possible to claim in this monograph to have reversed these findings.“

In ihrer Metaanalyse haben Scheerens/Bosker auch die Ergebnisse der von Rutter und Mortimore durchgeführten Sekundar- bzw. Grundschulstudien berücksichtigt. Diese beiden Studien gelten als prominente Nachweise der Lernwirksamkeit von Schulen aufgrund deren Schulethos. Der Aussage des Titel der Grundschulstudie („*School matters*“, Mortimore u.a. 1988) wird zwar von Scheerens/Bosker zugestimmt, das relative Ausmaß der Effekte der Schule wird hingegen deutlich anders bewertet. „The picture that emerges from this latter review is rather pessimistic from the perspective of school

effectiveness: the almost inevitable conclusion is that teacher effects dominate over school effects“ (Scheerens/Bosker 1997, S. 96).

Luyten (1994) hat eine vielbeachtete Zerlegung des Einzelschuleffektes publiziert, den er mit 15 Prozent in Relation zu 85 Prozent der Effekte von Schülermerkmalen beziffert. 25 Prozent des Schuleffekts (d.h. 0,25 von 15 Prozent = 4 Prozent der Varianz auf Schülerebene) kann als genuiner Institutionseffekt identifiziert werden, weitere 27 Prozent des Schuleffekts gehen auf die Interaktion von untersuchtem Schulfach und Schülerjahrgang zurück, 8 Prozent auf den Schülerjahrgang und 40 Prozent auf das Schulfach. Der Interaktionseffekt spiegelt vermutlich Lehrereffekte wider, da Lehrer in Schulen eher jahrgangs- und fachgebunden eingesetzt werden. Luyten (1994, S. 199) fasst seine Schlussfolgerung in einer Warnung zusammen: „The possibility that school effectiveness is actually an artefact and that effective schools are simply schools with a high percentage of effective teachers or departments should be taken seriously“.

Die Zerlegung der Schulvarianz in einen Kohorteneffekt (Schuljahrgang) und einen fächerspezifischen Effekt bietet Antworten auf die beiden kapitelingangs erwähnten Generalisierungsfragen. Wenn SER-Kennwerte – wie in QuaSUM und MARKUS sowie in PISA 2000 – in einer Schule nur für einen Schülerjahrgang erhoben werden, so muss die Frage gestellt werden, ob die Kennwerte eine hinreichende zeitliche Stabilität aufweisen und z.B. im Folgejahr ähnlich ausfallen würden. In der Längsschnittperspektive scheint dies der Fall zu sein, d.h. Wiederholungsuntersuchungen der gleichen Kohorten erbringen im Abstand von einem oder zwei Schuljahren ähnliche Ergebnisse. In der Querschnittperspektive, die schon aus Kostengründen größere Realisierungschancen hat, werden jedoch in einer Schule zeitgleich unterschiedliche Schülerjahrgänge untersucht. Die Konsistenz der SER-Kennwerte ist hier deutlich geringer (vgl. Ditton/Krecker 1995).

Raudenbush (1989) macht den Vorschlag, Schuleffekte in eine zeitstabile (höher reliable, durchschnittliche Effektivität) und in eine zeitvariable Komponente (weniger reliable, aktuelle Effektivität) zu zerlegen. Eine angemessene, allerdings keineswegs einfach interpretierbare Handhabung der zeitlichen Variabilität von schulischer Lernwirksamkeit wird von ihm darin gesehen, schulbezogene Wachstumskurven zu modellieren und diese als Schuleffekte zu betrachten. Schulentwicklungsprogramme sollten nachhaltige Wirkungen zeitigen und müssten deshalb an der zeitstabilen Komponente der Schuleffektivität ansetzen. Damit würde zugleich der Nachteil in Kauf genommen, dass aktuelle Initiativen zur Effektivitätssteigerung unberücksichtigt bleiben, was vermutlich auch bei Lehrern eine motivationspsychologisch ungünstige Strategie darstellt.

Fachübergreifende SER-Kennwerte (sog. composite scores) erbringen deutlich höhere Werte für den Schuleffekt als fachbezogene Kennwerte (vgl. Scheerens/Bosker 1997, S. 77). Dieses empirische Faktum illustriert nochmals das Dilemma: Rückmeldung der Schuleffizienz zeigt vor allem dann informative Unterschiede zwischen den Schulen, wenn Kennwerte verwendet werden, die innerhalb einer Schule schwer zu interpretieren sind, da eine Lehrerzuordnung kaum möglich ist. Dies kann auch wie folgt interpretiert werden: schulfachbezogene Kennwerte sind schon innerhalb einer Schule weniger konsistent und bieten deshalb kein einheitliches Bild „der“ Effektivität der Schule. Diesem Nachteil, der insbesondere die Anknüpfung von Schulentwicklungsstrategien erschwert, steht der Vorteil gegenüber, dass fachbezogene Kennwerte (departmental effectiveness) von einer Lehrergruppe (z.B. von der Fachkonferenz) oder von einzelnen Lehrern als ihre pädagogische Arbeit in direkter Weise betreffende Rückmeldung aufgefasst werden können.

Obendrein sind fächerübergreifende Vergleiche von SER-Kennwerten mit der Problematik belastet, dass den fachspezifischen Skalen zur Messung der Lernstände keine einheitliche bereichsbezogene Kompetenzdimension zugrunde liegen kann; Konzepte wie die ‚mathematisch-muttersprachliche Kompetenz‘ wirken artifiziell. Dass die Skalen trotz statistisch gesicherter einheitlicher Metrik vergleichbare Kompetenzen messen, ist keineswegs einfach zu begründen. Zudem belastet die unterschiedliche ‚Schwierigkeit‘ von Schulfächern den Vergleich von Skalenwerten (vgl. Newton 1997).

Im Primarschulbereich fallen SER-Kennwerte in geringerem Maße fachspezifisch aus als im Sekundarschulbereich. Die höhere fachübergreifende Konsistenz im Bereich der Grundschule begünstigt einerseits Schulentwicklungsstrategien. Andererseits unterrichten Grundschullehrer zumeist mit hohem Stundenanteil in ihrer Klasse und damit in der Jahrgangsstufe, sodass im Vergleich zu Sekundarschulen höhere Interaktionen zwischen Lehrer- und Jahrgangsstufeneffekten zu erwarten sind.

Zusammenfassend kann festgestellt werden, dass SER-Kennwerte nur eingeschränkt tauglich sind, die Lernwirksamkeit einer Schule so abzubilden, dass aufbauend auf diesen empirischen Informationen mittelfristige Initiativen der Schulentwicklung planbar sind. Besondere Probleme bereiten die Fakten, dass Sekundarschulen nicht einheitlich lernwirksam sind, sondern fachspezifisch unterschiedlich, und dass die Lernwirksamkeit von Primar- und Sekundarschulen in beträchtlichem Maße mit Lehrereffekten konfundiert ist, was nur mit hohem empirischen Aufwand aufschlüsselbar ist. Eine von Visscher (2001) veröffentlichte Befragung international führender SER-Experten zur Nutzung von Schuleffizienzwerten enthält deshalb konsequenter Weise ein Vielzahl von Kautelen und Warnungen.

5. Binnenschulisches Risikomanagement als komplexe Optimierungsstrategie für Schuleffektivität

In Deutschland werden derzeit zwei Typen von Schuleffektivitätsstudien realisiert. Den ersten Typus bilden Evaluationsstudien wie QuaSUM oder MARKUS, die auf einmalige Messungen beschränkt sind. Sie erbringen eine Abschätzung der Schuleffektivität für eine Klassenstufe und somit einen Jahrgang einer Schule in jeweils einem Schulfach (Mathematik). Die Generalisierbarkeit der Ergebnisse auf die Schuleffektivität in anderen Fächern ist nicht bekannt, ebenso wenig kann die Kohortenspezifität der Daten abgewogen werden.

Diese Einschränkungen sind nicht ein Problem der Studien, sondern eine Konsequenz des eingeschränkten Untersuchungsauftrags und somit im Wesentlichen auch Kostenfragen. Ein zentrales Problem von Einpunktmessungen berührt die Fairness der SER-Kennwertberechnung. Das fachspezifische Vorwissen stellt einen mächtigen Prädiktor für Schulleistungen dar und sollte im optimalen Design berücksichtigt und so erfasst werden, dass die Lernstände bei Eintritt in die Institution abgebildet werden. Spätere Messungen – in MARKUS wird der sicherlich elegante Kunstgriff verwendet, zeitgleich mit dem Lernstand auch den Stand des relevanten Vorwissens zu messen – sind immer mit dem Einwand konfrontiert, dass vorgängige, institutionseigene Lehrer- und Kohorteneffekte mitgemessen werden.

Aus der wissenschaftlichen Sicht sind die Anforderungen an faire, schulentwicklungsbezogene SER-Kennwerte beträchtlich. Für eine Abschätzung der stabilen Komponente im Verlauf der Effektivität einer Schule muss zumindest eine Dreipunktmessung realisiert werden: (1) Vorwissen bei Eintritt in die Institution, (2) Erstmessung und (3) Zweitmessung bei gleicher Lehrperson in gleichwohl hinreichendem Zeitabstand. Die Messpunkte (2) und (3) werden benötigt, um überhaupt einen Verlaufsgradienten bestimmen zu können; Messpunkt (1) dient zur Adjustierung bezüglich des Vorwissens. Um eine Konfundierung mit Fachlehrerwechseln möglichst weitgehend auszuschließen, sollten die Messpunkte (2) und (3) im gleichen Schuljahr liegen.

Den zweiten Typus von Schuleffektivitätsstudien bildet die LAU in Hamburg. In dieser Studie wird tatsächlich eine Schülerkohorte eines ganzen Bundeslandes zumindest dreifach untersucht und dies zumindest in allen Hauptfächern. Eine zeitnahe Adjustierung bezüglich des Vorwissens kann ebenso vorgenommen werden wie die Abschätzung von klassen- und schulspezifischen Verläufen der Lerneffektivität. Unklar bleibt allerdings die Übertragbarkeit auf andere Schülerkohorten, was auch als erhebliche Konfundierung der schulspezifischen SER-Kennwerte mit Lehrereffekten aufgefasst werden kann.

Als dritter Typus können schulbezogene Monitoring-Programme bezeichnet werden (vgl. Fitz-Gibbon 1996; Tymms 1999), an denen viele Schulen regelmäßig in diversen Fächern und Jahrgängen partizipieren. In Deutschland sind im Unterschied zu England oder den USA Studien dieser Art bislang nicht durchgeführt worden.

Die nachfolgenden Erörterungen beziehen sich auf den erstgenannten Typ von Schuleffizienzstudien (einmalige Messung) und diskutieren jene binnenschulischen und gelegentlich zwischenschulischen Prozesse, die vermutlich die mittelfristige Schuleffizienz maximieren und die zugleich in Einpunktmessungen zu einer Fehleinschätzung der Schuleffektivität führen. Rückmeldungen von SER-Kennwerten sind unter diesen Konstellationen fragwürdig, weil die Ergebnisverzerrung nicht ausgewiesen wird. Binnenschulisches Risikomanagement in der Lehrer- und Klassenzusammensetzung wird als eine latente gesellschaftliche Funktion von Schule aufgefasst, deren schulindividuelle Handhabung zur Lösung einer bildungspolitischen Problematik beiträgt. Die übergreifende Problemlage besteht darin, dass die drei zentralen Institutionswirkungen (Qualifikations-, Selektions- und Integrationsfunktion) nicht hinreichend ‚vereinbare Zielkriterien‘ darstellen. Manches spricht für die Vermutung, dass die bildungspolitisch gleichwohl intendierte mehrkriteriale Zielerreichung in erheblichem Maße von informellen Kompensationsmechanismen zwischen und innerhalb von Schulen profitiert, die kaum durch allgemeine Regelungen administrierbar zu sein scheinen und die gerade deshalb eine sehr beträchtliche Systemleistung der Schulen darstellen.

Die Analyse wird beschränkt auf eine spezifische schulinterne Managementaufgaben: die Zuordnung von Lehrern zu Klassen. Die bekannten Effektgrößen für zumeist indirekt erfasste Lehrer- und für gemessene Schülermerkmale fallen zumeist deutlich größer aus als die Effekte auf der Schulebene und können somit als hinreichende Begründung für die gewählte Analysestrategie angesehen werden.

Mit dem Begriff ‚Lehrerbeeinträchtigung‘ werden in unspezifischer Weise alle erheblichen Reduzierungen der pädagogischen und/oder didaktischen Leistung von Lehrern sowie erhebliche Fehlzeiten infolge von physischer und/oder psychischer Erkrankung bezeichnet. König (1999, S. 7) hält die „lenkende Anweisung“ von „Mitarbeitern mit geringer Kompetenz“ für eine in der Häufigkeit unterschätzte und in der Effizienz überschätzte Führungsfunktion von Schulleitern. Bessoth (1989, S. 108) vertritt die Hypothese, dass ungefähr 5 bis 10 Prozent der Lehrer der Gruppe „Nicht-akzeptable Leistungen von LehrerInnen mit keinen oder wenig Chancen der Leistungsrückgewinnung“ zuzuordnen sind.

Weiterhin wird davon ausgegangen, dass jeder Schulleiter jeden Lehrer seiner Schule bezüglich dieser Merkmale einschätzen muss, weil davon die

Qualität und Quantität des erteilten Unterrichts abhängig ist, deren Optimierung auf der Ebene der Schule zu den primären Dienstpflichten der Schulleitung gehört. Ob diese Einschätzungen hinreichend zuverlässig und zutreffend erfolgen, stellt eine heikle Frage dar, die für die nachfolgende Analyse bejaht werden soll. Es geht um den Extremfall der ‚guten Schulleitung‘, deren Risikomanagement und den Aussagewert von SER-Kennwerten aus Untersuchungen vom Typ 1.

Der Begriff des Risikomanagements trägt der von Diederich (1994, S. 250) sehr direkt formulierten Tatsache Rechnung, dass es „praktisch kaum eine Möglichkeit gibt, unfähige oder pflichtvergessene Lehrer aus dem Dienst zu entfernen“. Fachzeitschriften, die sich an Schulleiter wenden (z.B. *Schulmanagement* oder *Pädagogische Führung*), enthalten erstaunlicher Weise keine Beiträge, die sich der Problemstellung widmen, wie der unterrichtliche Einsatz insuffizienter Lehrer zu optimieren ist. Beiträge zum Führungsverhalten sind hingegen häufig darin zu finden und legen die Vermutung nahe, dass das Prinzip des Interessenausgleichs entweder als allparteiliche Nutzenmaximierung oder Schadensminimierung favorisiert wird. In diesem Sinne werden nachfolgend mögliche Varianten des Risikomanagements ausformuliert.

5.1 Klassengröße – Lehrerkompetenz

Ein vermutlich häufig praktiziertes Risikomanagement besteht darin, den beeinträchtigten Lehrer in einer Parallelklasse, die deutlich weniger Schüler hat oder im Verlauf des Schuljahrs durch Abgänge bzw. die Andersplatzierung von Neuzugängen haben wird, einzusetzen (klassengrößenbezogene Kompensation). Da die Schätzgenauigkeit von SER-Kennwerten in direkter Beziehung zur Klassengröße steht, ergibt sich folgende Konsequenz: Der Schulklasseneffekt der Kompensationsstrategie wird deutlich ungenauer geschätzt und zugleich näher an den Mittelwert der Parallelklasseneffekte herangezogen (Schrumpfung in HLM, vgl. Arnold 1999b, S. 225). Somit tritt ein echter und ein artifizieller Kompensationseffekt ein; das Rechenmodell bessert das Ausmaß der Effektivitätsminderung nach.

Aus dem methodisch gleichen Grund können SER-Kennwerte auch zu doppelten Unterschätzungen führen. Eine schulbezogene Optimierungsstrategie besteht darin, den Unterricht in einer Klasse, in die ein oder mehrere behinderte Schüler integriert werden, nur pädagogisch besonders kompetenten Lehrern zu übergeben. Viele Ergebnisse der Integrationsforschung weisen darauf hin, dass die nicht-behinderten Schüler auch im fachlichen Lernen erheblich von der integrativen Beschulung profitieren. Wenn jedoch Integrationsklassen – aus besten Gründen – deutlich weniger Schüler als die Paral-

lellklassen haben, dann tritt für die Berechnung von SER-Kennwerten das oben beschriebene Phänomen in gegenläufiger Richtung auf: die SER-Kennwerte werden nach unten in Richtung der vermutlich weniger effektiven Parallelklassen korrigiert.

Das sich hier zeigende methodische Problem kann in allgemeiner Form so ausgedrückt werden: korrekte statistische Kalkulation führt in Schulen mit sehr unterschiedlichen Klassengrößen dazu, dass der Schulkennwert von den Ergebnissen der größeren Klassen stärker beeinflusst wird. Dies kann unter pädagogischen Gesichtspunkten jedoch unfair erscheinen und auf der Klassenebene zu Fehleinschätzungen führen. Burstein (1986, S. 59) hat in Abwägung der statistischen Gegenargumente vorgeschlagen, die Gewichtung der Kennwerte mit der Schülerzahl und damit ein zentrales Qualitätsmerkmal von HLM aufzugeben.

5.2 Klassenkomposition – Lehrerkompetenz

Ein weiteres Risikomanagement besteht darin, den beeinträchtigten Lehrer in der Parallelklasse mit den fähigsten Schülern einzusetzen (klassenfähigkeitsbezogene Kompensation). Diese Strategie kann sich auf Ergebnisse der Schuleffizienzforschung berufen: schwächere Schüler profitieren stärker von gutem Unterricht (vgl. Scheerens/Bosker 1997, S. 96). In diesem Fall werden damit die Effizienzunterschiede zwischen den Parallelklassen nivelliert und in den rückgemeldeten Schul- sowie Klasseneffizienzwerten zeigt sich eher das Gegenteil von den realen Verhältnissen: eine relativ homogen effektive Schule, die eine ideale Voraussetzung für schulbezogene Entwicklungsstrategien böte. Als strukturell ähnliche Konstellation tritt vermutlich noch häufiger auf, dass Schüler mit Disziplinarmaßnahmen (Klassen- oder Schulverweis) vorzugsweise in die Klasse eines hochfähigen Lehrers gegeben werden, dessen klassenbezogene Unterrichtseffektivität aufgrund der intensiven Beschäftigung mit dem Problemschüler über längere Zeit reduziert werden kann. Auch diese binnenschulische Optimierungsmaßnahme führt zu einer eher verzerrenden Homogenisierung der SER-Kennwerte.

5.3 Unterrichtsfach – Lehrerkompetenz

Ein in der Grund- bzw. Hauptschule häufig praktiziertes Optimierungs- bzw. Risikomanagement besteht darin, den Klassenlehrer mit möglichst vielen Stunden in seiner Klasse einzusetzen, um ein hohes Maß an personbezogener und konstanter pädagogischer Zuwendung zu realisieren. In Kauf genommen

wird ein nicht unbeträchtliches Ausmaß an fachfremdem Unterricht, den Lehrer vermutlich mit eher geringerer Lernwirksamkeit erteilen (klassenlehrerbezogene Kompensation). Diese Konstellation kann mit bildungspolitischen Mitteln nicht sinnvoll korrigiert werden; ‚Allround-Lehrer‘ sind vermutlich nicht ausbildbar. Auf der Ebene der Einzelschule werden gleichwohl mit spezifisch geringeren SER-Kennwerten Ausgleichsprozesse hergestellt.

Fachbezogene Schulprogrammschwerpunkte oder Schulprofilbildung können zu systematischen Verzerrungen von SER-Kennwerten führen. Wenn – wie schon aus logischen Gründen erwartbar – Schwerpunktsetzungen eher in den so genannten Nebenfächern erfolgen, werden Schulleitungen zur Absicherung der Innovation vorzugsweise hochfähige Lehrer in diesen Bereichen einsetzen und so die hauptfachbezogenen Kompensationsmöglichkeiten für den Einsatz von beeinträchtigten Lehrern vermindern. Damit wird schulbezogen der Anteil hochfähiger Lehrer am Hauptfachunterricht reduziert. Wenn wie üblich SER-Kennwerte primär für hauptfachbezogene Schulleistungen erhoben werden, scheint eine solche Schulprogrammstrategie nicht opportun zu sein.

5.4 *Lehrerschaft der Klasse – Lehrerkompetenz*

Eine Variante des Risikomanagements kann darin bestehen, einen beeinträchtigten Lehrer vorzugsweise in Klassen mit hochfähigen Fachlehrern einzusetzen (klassenlehrerschaftsbezogene Kompensation). Da zumeist nicht allzu viele Klassen diese Lehrerkomposition aufweisen, wird der Unterrichtsanteil des beeinträchtigten Lehrers mehr als ein Fach umfassen. SER-Kennwerte, die just für die von jenem Lehrer unterrichteten Fächer gerechnet werden, unterschätzen die in den anderen Fächern für diese Klasse erreichte Lerneffektivität. Vermutlich wird dieses schulbezogen durchaus rationale Risikomanagement jedoch selten praktiziert, weil Schulleiter den Protest engagierter Klassenelternschaften fürchten müssen, die sich durch den Eindruck hoher Lerneffizienz in den anderen Schulfächern ermutigt fühlen können, Lehrerwechsel in dem Fach oder den Fächern zu verlangen, die von einem beeinträchtigten Lehrer unterrichtet werden. Insofern werden Schulleiter vermutlich eher dazu neigen, besonders beeinträchtigte Lehrer in möglichst vielen Klassen einzusetzen (Zwischenklassengesamtkompensation). Ein solcher Unterrichtseinsatz ist wiederum für den betreffenden Lehrer hochgradig belastend und vermindert zusätzlich seine eh reduzierte Kompetenz. Eine jahrgangsübergreifende Risikoverteilung kann hier jedoch als die optimale Strategie zum Erreichen hoher SER-Kennwerte sein, die in Studien des Typ 1 selten mehrere Jahrgänge umfassen.

5.5 *Lehrerteams – Lehrerkompetenz*

Vor allem in Gesamtschulen ist das von der IGS Göttingen Geismar entwickelte Team-Kleingruppen-Modell (vgl. Brandt/Schlömerkemper 1985) verbreitet. Die Lerngruppen jeweils einer Jahrgangsstufe werden von einem über Jahre weitgehend gleichen Lehrerteam unterrichtet. Wenn, was sozialpsychologisch durchaus wahrscheinlich ist, diese Gruppierungen auch aufgrund von Berufskompetenz gebildet werden, ist die Folge unausweichlich, dass in diesen Schulen beträchtliche Kohortenunterschiede in der Schuleffektivität auftreten. Jahrgangsbasierte SER-Kennwerte sind dann entweder positiv oder negativ verzerrt, wenn es überhaupt so etwas geben sollte wie die jahrgangsübergreifende Effizienz einer so organisierten Schule; in der Terminologie von SER handelt es sich um eine Sonderform von departmental effectiveness. Ob diese Art von zeitlich hochstabiler Teambildung zu einer Optimierung der einzelschulbezogenen Lerneffektivität führt, ist aus theoretischer Sicht kaum vorherzusagen. Da die empirische Untersuchung dieser hochinteressanten Fragestellung einen sehr großen Aufwand erfordert, wird auch die faktenbezogene Klärung nicht leicht zu erhalten sein. Das TKM erschwert die SER-Kennwertnutzung auch aus einem bereits erörterten Grund: geringe Schülerzahlen nivellieren Effizienzunterschiede.

Für Lehrertandems, die bei Doppelbesetzung in Integrationsklassen gebildet werden, ist durchweg eine Effektivitätssteigerung zu erwarten. Allerdings beinhalten diese wie alle klassenbezogenen Lehrerdyaden auch ein besonderes Risiko: Wenn zwischenmenschliche Konflikte hier auftreten, ist infolge der beeinträchtigten Kooperation eine Minderung der Unterrichtsqualität kaum zu vermeiden. Werden SER-Kennwerte unter diesen Bedingungen erhoben, bilden diese in erheblichem Maße Konstrukt fremdes, d.h. passagere psychosoziale Belastungen einzelner Lehrer ab. Die unterstützende Beratung oder – nach deren Scheitern – die schuleffizienzorientierte Neukonstituierung von Lehrerteams gehört vermutlich zu den besonders schwierigen Aufgaben der Schulleitung.

5.6 *Regionale Schulökologie*

Aus der Schuleffizienzforschung liegen unterschiedliche Befunde zu Kompositionseffekten auf der Schulebene vor (geringe Effekte: vgl. Sammons u.a. 1996; größere Effekte: vgl. Nuttall u.a. 1989). Klassenbezogene Kompositionseffekte können beträchtlich sein und müssten aus der Perspektive der Schulleitung entweder binnenschulisch egalisiert oder für das Risikomanagement kompensatorisch genutzt werden. Wenn in einzelne und nicht alle Parallel-

klassen von Sekundarschulen ganze oder hinreichend große Teile von Grundschulklassen übernommen werden, um Lerngruppenkontinuität zu wahren, werden Kompositionseffekte längerfristig klassenbezogen fortgeführt. Es lassen sich auch didaktische Gründe für die Vermutung anführen, dass diese parallelklassenungleichen Effekte nicht hinreichend auf der Einzelschülerebene (Vorwissen) und auf der Klassenebene (Abweichung des Klassenmittelwerts vom Schulmittelwert) adjustiert werden können und somit verzerrte SER-Kennwerte resultieren, denn unterrichtserleichternde Bedingungen können unabhängig vom Leistungsniveau z.B. dadurch eintreten, dass die erhöhte Kohäsion in der Teilgruppe ein hohes Maß an routinierter Gruppenarbeit ermöglicht.

Abschließend sollen zwei besonders heikle, informelle Risikostrategien angesprochen werden. Nicht selten wird in einer Schulregion bekannt, an welcher Schule besonders große Toleranz und/oder hohes pädagogisches Engagement für Schüler mit disziplinarischem Schulverweis besteht. Die Schulaufsicht, aber auch die Beratungseinrichtungen und gelegentlich auch die Eltern dieser von schulischem Scheitern bedrohten Schüler werden diese Schule aus besten Gründen um Aufnahme bitten. Im Modell der SER-Kennwerte ist jedoch keine Adjustierung für die effektivitätsmindernde Wirkung der Aufnahme einzelner verhaltensauffälliger Schüler vorgesehen, so dass die pädagogisch besonders qualifizierte und für das bessere Funktionieren der relegierenden Nachbarschulen hochwertige Schule benachteiligt wird.

Ähnliches gilt auf der Ebene der Lehrerzuweisung. Insbesondere in überschaubaren Schulverwaltungsregionen z.B. von Stadtstaaten wird – zumindest hinter vorgehaltener Hand – über Einzelfälle von beeinträchtigten Lehrern als ‚Wanderpokale‘ oder ‚Blaupapier‘ geredet. Was auch immer an Vorurteilshaftigkeit sich hier artikulieren mag – Tatsache ist, dass es Fälle von pädagogisch insuffizienten Lehrern gibt und dass diese Fälle zumeist nicht aus der unterrichtlichen Tätigkeit genommen werden. Bessoth (1989, S. 124) meint, „2–4% ‚schwarze Schafe‘“ ausmachen zu können, „die in den Schulen beachtliche pädagogische Flurschäden anrichten“. Wenn alle Schulen gleichmäßig mit der Aufnahme oder Weiterbeschäftigung dieser Lehrer belastet würden, ergäbe sich kein grundsätzliches Fairnessproblem für die Rechnung von SER-Kennwerten. Wenn jedoch, was anzunehmen ist, diese Lehrer nicht einheitlich in bestimmten Klassenstufen eingesetzt werden, dann können SER-Kennwerte im Falle einer einzelnen Schule zu erheblichen Ergebnisverzerrungen führen.

Dies gilt umso mehr, wenn es einzelne Schulen gibt, von denen bekannt ist, dass deren Kollegium und Schulleitung sich für die berufliche Entwicklung insuffizienter Lehrer besonders engagieren – oder wenn aufgrund an-

derweiteriger Abhängigkeiten (z.B. Personalknappheit) Schulleitungen die Aufnahme eines aufgrund von Insuffizienz versetzten Lehrers nicht ablehnen können. Personalüberhang kann gleichfalls ein solcher Grund sein; die Aufnahme des insuffizienten Lehrers wird nicht als Kapazitätserweiterung gerechnet. Ob diese Konstellation zu einer Erhöhung der Effizienz auf der Schulebene führt oder einen Ausgleich von Nach- und Vorteilen bringt oder negative Wirkung zeitigt, ist schwerlich abzuschätzen. Auch dieses Beispiel kann als Beleg für die eingangs formulierte These gelten, dass eine systembezogene Optimierung von Schuleffizienz offensichtlich auf subtile einzelschulbezogene Prozesse angewiesen ist, die Bildungspolitik nicht erfassen kann und zum Vorteil einzelner Lehrer vielleicht auch mit Diskretion behandeln sollte. Gleichwohl wäre eine interessante Fragestellung der Schuleffizienzforschung, die kollegiale Rehabilitationsleistung von Schulen und Schulsystemen zu erfassen.

6. Perspektiven der Nutzung von Schuleffizienzkennwerten

Schulleistungsstudien sind ein hochwertiges Instrument zum Monitoring der tatsächlich erreichten Lernstände und dienen damit sowohl der Schuladministration als auch den Lehrern als eine besondere Gelegenheit, das Ausmaß der lehrplanbezogenen Zielerreichung abzuschätzen. Für den Part der Lehrer resultiert daraus die Gelegenheit zur Kalibrierung ihrer Leistungsbeurteilung. Schuladministrationen könnten die Erreichbarkeit von Lehrplanvorgaben empirisch abschätzen. Schuleffizienzmessung mit einzelschulbezogener Rückmeldung von SER-Kennwerten ist zumindest mittelfristig nur im Setting von flächendeckenden Schulleistungsstudien zu realisieren und insofern an ein hohes Maß von Kooperationsbereitschaft der Lehrer gebunden, das bekanntermaßen durch unfaire Schul- und Klassenvergleiche, vielleicht sogar durch jedwede Form der Veröffentlichung von Leistungskennwerten reduziert wird.

SER-Kennwerte gehen über die Mitteilung von Lernständen (d.h. von rohwerthasierten Parametern) deutlich hinaus und berücksichtigen den Kontext schulischer Leistungserbringung. Das Konstrukt Schuleffizienz besitzt jedoch eine nur begrenzte Generalisierbarkeit. Mit den in Deutschland bislang realisierten Studien, die überwiegend als Einzelpunkt- bzw. Einkohortenmessungen angelegt sind, wird auf der Ebene der ganzen Schule so wenig Information erhoben, dass damit keine hinreichend generalisierbare Grundlage für deren mittelfristige Entwicklungsplanung erschlossen wird. Allerdings bieten alle bekannten anderen Verfahren – insbesondere jene der schulinternen Evaluation – Datenlagen von methodisch deutlich ungünstigerer Quali-

tät, sodass SER-Kennwerte gewiss zu den *relativ* besten Entscheidungsparametern gehören.

Für die traditionellen pädagogischen Arbeitsstrukturen innerhalb einer Schule, d.h. für die vom Untersuchungsdesign betroffenen einzelnen Fachlehrer und ebenso für die betroffenen Klassen-, Stufen- und Fachkonferenzen bieten SER-Kennwerte faire Rückmeldungen zu einem wichtigen Teilaspekt des beruflichen Handelns. Als Verbesserungsmaßnahme wirkt vermutlich bereits die Rückmeldung selbst, und ansonsten die Nutzung der Standardinstrumentarien wie Beratung, Supervision und Lehrerfortbildung.

Die empirischen Fakten der Schuleffizienzforschung erbringen eine klare Evidenz, was die relative Größenordnung von Varianzquellen im Schulsystem betrifft. Für Deutschland muss festgestellt werden, dass die Schulform eine sehr erhebliche Varianzquelle bildet und insofern Bildungspolitik als systembezogene Effizienzsteigerung wieder in den Blick gerät. Eine gleichfalls relativ große Varianzquelle liegt auf der Ebene der Klassen: im Vergleich zu den geringeren Schuleffekten sind größere Lehrer- und Klasseneffekte auffindbar. Angesichts dieser Befundlage liegt die Schlussfolgerung nahe, dass das Konzept der Schulreform als Entwicklung der Einzelschule an einer zweitrangigen Varianzquelle ansetzt, was nicht gegen eine solche Strategie spricht, sondern gegen die Hoffnung, deren Planung auf der Basis von Schuleffizienzdaten zu betreiben.

Für die bildungspolitische Systemplanung erbringen SER-Kennwerte eine äußerst wichtige, derzeit wohl aber eher enttäuschende Information. Nur über Effizienzstudien dieser Art lässt sich abschätzen, wie die soziale Gerechtigkeit des Schulsystems entwickelt wird. Nehmen soziale Disparitäten zu, stagniert deren Verringerung oder wird kontinuierlich ein Mehr an sozialer Gerechtigkeit erreicht?

Dabei sollte nicht übersehen werden, dass diese gesellschaftliche Zielerreichung vermutlich in beträchtlichem Maße auch durch subtile Prozesse eines nicht administrierten Risikomanagements auf der Ebene der Einzelschule gewährleistet wird. Dieser Effizienzbeitrag der Einzelschule ist durch Modellierung kaum in die Berechnung von SER-Kennwerten einzubringen, sodass je nach Risikolage in der einzelnen Schule Ergebnisverzerrungen eintreten können. Manches spricht für die bereits zitierte These von Luyten, die in pointierter Form auch so lauten könnte: Effektive Schulen im Sinne von SER-Kennwerten sind einfach jene, die einen sehr geringen Prozentsatz beeinträchtigter Lehrer und einen ebenso geringen Anteil verhaltensgestörter Schüler haben *und* halten können. Jedoch: diese Schulen kann es nur geben, weil andere Schulen die unvermeidbaren Rehabilitationsleistungen erbringen.

Literatur

- Arnold, K.-H. (1999a): Fairness bei Schulsystemvergleichen: Diagnostische Konsequenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation. Münster: Waxmann.
- Arnold, K.-H. (1999b): Schulen im Vergleich. Probleme des Ranking und Chancen eines Monitoring. In: *Die Deutsche Schule* 91, S. 218-231.
- Arnold, K.-H. (2001): Schulleistungsstudien und soziale Gerechtigkeit. In: *Zeitschrift für Pädagogik* 47, S. 161-177.
- Bastian, J. (1997): Pädagogische Schulentwicklung. Von der Unterrichtsreform zur Entwicklung der Einzelschule. In: *Pädagogik* 49, H. 2, S. 6-11.
- Baumert, J. (1997): Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). Dokumentation Bd. 1. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J. (2001): Internationale Schulleistungsvergleiche. In: Rost, D.H. (Hrsg.): *Handwörterbuch Pädagogische Psychologie*. Weinheim: Psychologie Verlags Union, S. 294-303.
- Bessoth, R. (1989): Regelbeurteilung zur Leistungsfeststellung (Lerneinheit 35.04). In: Bessoth, R./Schmidt, H. (Hrsg.): *Schulleitung – ein Lernsystem*. Bd. 3a: Personal – Schulklima. Neuwied: Luchterhand, S. 1-144.
- Brandt, H./Schlömerkemper, J. (1985): Kommunikative Lerndiagnose: Konzept und Wirklichkeit des Lernentwicklungsberichts im Team-Kleingruppen-Modell. In: *Zeitschrift für Pädagogik* 31, S. 201-219.
- Burstein, L. (1986): Erklärungsmodelle mit Zwischen- und Innerklassenregressionen: Grundlegende Konzepte und ein Beispiel. In: Saldern, M. v. (Hrsg.): *Mehrebenenanalyse*. Weinheim: Psychologie Verlags Union, S. 44-70.
- Cheung, K.C./Keeves, J.P. (1990): Testing multilevel product-process networks. In: *International Journal of Educational Research* 14, S. 307-316.
- Coleman, J.S./Campbell, E.Q./Hobson, C.J./McPartland, J. (Hrsg.) (1966): *Equality of educational opportunity*. Washington, D.C.: US Government Printing Office.
- Diederich, J. (1994): Der Lehrer. In: Lenzen, D. (Hrsg.): *Erziehungswissenschaft. Ein Grundkurs*. Reinbek: Rowohlt, S. 228-252.
- Ditton, H. (1995): Ungleichheitsforschung. In: Rolff, H.-G. (Hrsg.) *Zukunftsfelder der Schulforschung*. Weinheim: Deutscher Studien Verlag, S. 89-124.
- Ditton, H. (1998): Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells. Weinheim: Juventa.
- Ditton, H./Krecker, L. (1995): Qualität von Schule und Unterricht. Empirische Befunde zu Fragestellungen und Aufgaben der Forschung. In: *Zeitschrift für Pädagogik* 41, S. 507-528.
- Fend, H. (1982): *Gesamtschule im Vergleich*. Weinheim: Beltz.
- Fitz-Gibbon, C.T. (1996): *Monitoring education. Indicators, quality and effectiveness*. London: Cassell.
- Giesecke, H. (1995): Wozu ist die Schule da? In: *Neue Sammlung* 35, S. 93-104.
- Goldstein, H. (1995): *Multilevel statistical models*. London: Arnold.
- Goldstein, H./Spiegelhalter, D.J. (1996): League tables and their limitations: Statistical issues in comparisons of institutional performance. In: *Journal of the Royal Statistical Society A* 159, S. 385-443.
- Helmke, A. (2001): Internationale Schulleistungsvergleichsforschung. Schlüsselprobleme und Perspektiven. In: *Zeitschrift für Pädagogik* 47, S. 155-160.

- Helmke, A./Jäger, R.S. (Hrsg.) (2001): Die Studie MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext. Grundlagen und Perspektiven. Landau: Verlag Empirische Pädagogik.
- Hopkins, D. (1996): Towards a theory for school improvement. In: Gray, J./Reynold, D./Fitz-Gibbon, C. (Hrsg.): Merging traditions: The Future of research on school effectiveness and school improvement. London: Cassell, S. 30-50.
- Hosenfeld, I./Helmke, A./Ridder, A./Schrader, F.-W. (2001): Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. In: Empirische Pädagogik 15, S. 513-534.
- Klippert, H. (1999): Unterrichtsentwicklung konkret. Hilfen zur Entwicklung einer schülerzentrierten Lernkultur. In: Pädagogik 51, H. 2, S. 12-14.
- Köller, O./Baumert, J. (1999): Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichbarer Standards. In: Zeitschrift für Erziehungswissenschaft 3, S. 384-430.
- König, U. (1999): Situations- und personengemessene Mitarbeiterführung (C 1.4). In: Buchen, H./Horster, L./Rolf, H.-G. (Hrsg.): Schulleitung und Schulentwicklung. Stuttgart: Raabe/Klett, S. 1-19.
- Lehmann, R./Gänsfuß, R. /Peek, R. (1999): Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7 –. Bericht über die Untersuchung im September 1998. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.
- Lehmann, R.H./Peek, R./Gänsfuß, R./Lutkat, S. (2000): QuaSUM: Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg. Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Luyten, H. (1994): Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. In: International Journal of Educational Research 21, S. 197-216.
- Luyten, H./Jong, R.de (1998). Parallel classes: Differences and similarities. Teacher effects and school effects in secondary schools. School Effectiveness and School Improvement, 9, H. 4, 437-473.
- Maritzen, N. (2001): Schulleistungsforschung und Schulentwicklung, zwei Seiten einer Medaille. In: Journal für Schulentwicklung 5, H. 2, S. 46-54.
- Mortimore, P./Sammons, P./Stoll, L./Lewis, D. (1988): School matters. London: Chapman.
- Newton, P.E. (1997): Measuring comparability of standards between subjects: Why our statistical techniques do not make the grade. In: British Educational Research Journal 23, S. 433-450.
- Nuttall, D.L./Goldstein, H./Prosser, R./Rasbash, J. (1989): Differential school effectiveness. In: International Journal of Educational Research 13, S. 769-776.
- Preuss-Lausitz, U. (1997): Soziale Ungleichheit, Integration und Schulentwicklung. In: Zeitschrift für Pädagogik 43, S. 583-596.
- Raudenbush, S.W. (1989): The analysis of longitudinal, multilevel data. In: International Journal of Educational Research 13, S. 721-740.
- Rolf, H.-G. (1991): Schulentwicklung als Entwicklung von Einzelschulen? Theorien und Indikatoren von Entwicklungsprozessen. In: Zeitschrift für Pädagogik 23, S. 865-886.
- Sammons, P./Mortimore, P. /Thomas, S. (1996): Do schools perform consistently across outcomes and areas? In: Gray, J./Reynold, D./Fitz-Gibbon, C. (Hrsg.): Merging traditions: The future of research on school effectiveness and school improvement. London: Cassell, S. 3-29.

- Scheerens, J./Bosker, R.J. (1997): The foundations of educational effectiveness. Oxford: Pergamon.
- Tymms, P. (1999): Baseline assessment and monitoring in primary Schools: Achievements, attitudes and value-added indicators. London: David Fulton.
- Visscher, A.J. (2001): Public school performance indicators: Problems and recommendations. In: Studies in Educational Evaluation 27, S. 199-214.
- Willms, J.D. (1992): Monitoring school performance: A guide for educators. London: Falmer.

Abstract: *Within the setting of broadly designed studies on school achievement, School Effectiveness Research (SER) allows to define in a fair manner the learning effectiveness of the individual school. However, due to the design of most of these studies, the generalizability of the results on the level of the individual school is restricted so much so that it seems hardly possible to plan and evaluate, on the basis of the empirical evidence, a school development focussing on the individual school as a unit of action. On the other hand, the teacher- and class-related as well as the age-group- and subject-related feedback on learning effectiveness offers valuable information for the traditional forms of advice, supervision, and further education for teachers below the school level. Beyond the school level, i.e. on the level of the school system, SER-data provide highly significant evaluation data on the development of social justice. Not enough attention is paid to the fact that, on the level of the individual school, we are dealing with subtle processes of risk management which, on the one hand, maintain the effectiveness of the entire system which, on the other hand, however, can hardly be captured by SER indicators.*

Anschrift des Autors:

Prof. Dr. Karl-Heinz Arnold, Technische Universität Berlin, Fakultät I, Institut für Erziehungswissenschaft, Pädagogische Psychologie, Franklinstr. 28/29 (FR4-3), 10587 Berlin.